

## A NEW EXPERIMENTAL TECHNIQUE FOR GATHERING SIMILARITY RATINGS FOR SOUNDS

*Eoin Brazil, Mikael Fernström*

Interaction Design Centre  
Department of Computer Science and  
Information Systems,  
University of Limerick, Ireland  
eoin.brazil@ul.ie

*Laura Ottaviani*

Università di Verona  
Dipartimento Scientifico e Tecnologico,  
Strada Le Grazie 15,  
I-37134 Verona, Italy  
ottaviani@sci.univr.it

### ABSTRACT

Multidimensional scaling techniques (MDS) are a vibrant area of research with much development and advancement in last few decades. In this paper we focus on an interactive 2-dimensional interface for creating a similarity space for audio. Classical MDS techniques can place great demands upon participants overwhelming their sensory and cognitive abilities to make choices across such datasets using only pairwise comparison. We outline the results of an investigation to obtain multi-dimensional similarity ratings for a mixed sound collection containing both recorded sounds and synthesised sounds, and we describe briefly a computer application for collecting such data that reduces task demands by use of a 2-dimensional visual display with multistream audio for organising sound similarity and returns the results in the context of the whole stimuli set.

### 1. Introduction

The creation of a multi-dimensional similarity space is a difficult and time-consuming problem. The typical method for creating such a space is to use pairwise comparison technique but this is a time consuming task, as each sound must be compared with all other sounds within a dataset. The number of comparisons required by each participant is shown in the following equation:

$$C = N(N-1)/2$$

where C is the number of comparisons required and N is the number of stimuli within the dataset. Taking an example of 60 audio stimuli a participant would have to make 1770 judgments to complete the similarity-rating task. The number of comparisons in this type of task requires a high degree of focus from the participant with corresponding levels of fatigue. As the number of comparisons grows it becomes increasing more questionable as to whether participants can maintain a consistent judgement across comparisons. Addressing the issues of fatigue and consistency we further developed an existing interface [1] for use as a multi-dimensional similarity rating application.

The Sounding Object Project (SOB) has pioneered several recent attempts to couple physical simulations to efficient synthesis techniques. The development of parametric models or Sound Objects has lead to the use of the Sonic

Browser for testing and validating the sounds produced by these sound models [2]. The Sonic Browser was developed as tool for accessing sounds or collections of sounds using sound spatialization and context-overview visualization techniques. The Sonic Browser uses multiple stream audio activated by cursor/aura-over-icons, representing sound files. The aura is the receiver of information shown as a grey circle surrounding the cursor in Figure 1, the cursor is surrounded by the aura. All sonic objects within the aura play simultaneously, which facilitates our ability to switch our attention between different sounds in the auditory scene, making use of the “cocktail party effect” [3]. The aura is user controlled and can be reduced or increased in size, by reducing the aura a user can “zoom in” on a particular sound/s. The interface was extended in the context of this work as an interface for obtaining multi-dimensional similarity ratings. A sound model is a synthetic caricature of a sound, where the salient features of the sound have been identified and can be parametrically synthesised. The perceptual scaling experiment yields data that can be used to inform and calibrate the features of these sound models. The Sonic Browser removes the constraints of pairwise comparisons, which allows participants to create richer sets of perceptual and cognitive criteria in their judgments. One of the constraints of randomised presentation in pairwise comparisons makes it impossible for participants to exert control over the order of comparisons or to adjust their similarity rating for a previously presented pair in light of a new criterion that he or she may have generated on the basis of a later comparison. Another constraint is the inability of pairwise comparison to allow the simultaneous comparison of multiple (three or more) stimuli, which may have similar attributes in common as the listener may miss the similarity when comparing only two sounds at the same time. An important issue to be aware of is that due to the two-dimensional constraint imposed by the computer screen, continuous similarity ratings made using the Sonic Browser are limited to two dimensions.

In the Sonic Browser a central idea is to map sound clips to aural and/or visual objects with properties that convey information about the sound clips and use the objects in order to create browsing spaces. The foundations of our design approach for the Sonic Browser are based on the principles of direct manipulation and interactive visualisation interfaces proposed by Shneiderman [4]. The three primary facets of this foundation are “overview first, zoom and filter, then detail on demand”. We avoid presenting a detailed introduction on the Sonic Browser<sup>1</sup>

as this has been discussed by Brazil et al [1]. The sound object models and the Sonic Browser are available for download from the SOb website at <http://www.soundobject.org>. Pure Data (PD) is a software system for live musical and multimedia performances. The sound object models created by the SOb project used PD for the synthesized sounds. PD files are also commonly referred to as patches or modules.

New approaches have been developed to enable the creation and analysis of multi-dimensional similarity space more quickly, e.g. Scavone et al application, the Sonic Mapper [5]. This application allows for the creation and analysis of multi-dimensional similarity spaces using single stream browsing on the Linux platform.

### 1.1. Scope of this study

In this study we have gathered similarity ratings from users who rated sound files of bouncing sounds using the Sonic Browser with multiple stream audio activated by cursor/aura-over-icons, representing sound files within a 2-D plot. We used several types of sound files consisting of both real recordings and sound models (synthesised sounds). The sound files represent balls bouncing events whether synthesised or recorded. The aim was to provide fast and direct access to the sounds, so users can easily rate a number of sounds in parallel and within the context of the whole collection of stimuli. The tight coupling between the visual and auditory information allows users to get a good spatial idea of what sounds are present and allows for easy scaling of these sounds within a 2-dimensional visual display.

## 2. Tasks

The scenario for this study is to rate bouncing sounds within a 2-dimensional scale. Users were requested to listen to the stimuli and to drag-and-drop them according to their judgements within a specified 2-D scale. The specified scale was a 2-D plot with the perceived size of the dropped object on the X-axis and the perceived height of the object drop on the Y-axis. After the scaling task was complete, participants were asked to judge the realism of the sounds used within the experiment by 'tagging' those sounds they felt were unrealistic. At the end of each session, a questionnaire was presented to the participants in order to gain an insight into their feelings about the performed tasks. The experimental data collection involved two techniques. First, data logging was collected by the application for the object positioning in the 2-D space. Secondly, the user's actions were captured on video, which was then analysed.

The dataset used in this experiment included 6 real sounds and 12 synthetic, 6 of which are designed with the PD-modules modelling impact interactions of two modal resonators [6], simplified returning only one mode, while the other 6 with the PD-modules modelling impact interactions of two modal resonators as well as the dropping event. The recorded sounds were produced by 3 steel balls, weighting 6, 12 and 24 g, and falling on a wood board of 1500 x 500 x 20 mm from a height of 10, 20 and 40 cm, respectively, by positioning the microphone at 3 different distances: 20 - 40 - 80 cm, respectively. The recording of these sounds was performed using a MKH20 Sennheiser microphone and at a sampling rate of 44.1kHz.

These stimuli were used in previous experiments conducted by the SOb project on the perception of impact sounds [7]. In this study, Burro found the relationship between the physical quantities of weight, distance and height and the relative perceptual quantities. He argues that manipulating one of the physical parameters affects more than one of the perceptual quantities. In our experiment we kept the distance constant ( $d=80$  cm) but changed the height of the drop of the object. The sound level of the Sonic Browser was controlled in this experiment so that loudness was related to distance from the aura centre.

## 3. Users

Five postgraduate students were recruited to partake in this study. All the participants referred to have musical training in average of 8 years, with a minimum of 6 and a maximum of 10 years. Another four postgraduate students were recruited for the initial pilot stage of this study.

## 4. Experiment

This experiment was an exploratory experiment to both further our understanding of gathering similarity ratings for sounds and to gather participant perceptions of the sounds, especially the synthesised sounds. We examined differences in the scaling of the real and synthesised sounds and participants observations of the task. In particular, we collected formative data relevant to the creation a similarity scaling using a 2-D plot for the particular stimuli collection. The participants were asked to complete two tasks, the scaling task and a realism judgement task of the stimuli used within the experiment. In each specific task, the users were allowed to move the cursor around freely in the GUI trying to find target sounds.

### 4.1. Experiment Design

Using a within-subjects design, subjects sorted a number of sound objects against the two properties on an x-y axis; these objects were initially presented in a random order, the task order was not counter-balanced as performing the scaling task first ensured a knowledge of the entire dataset prior to carrying out the realism task.

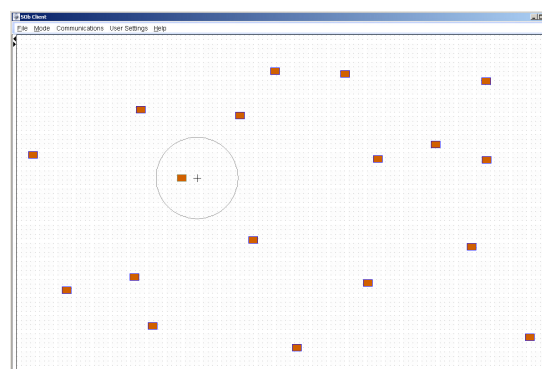


Figure 1: The Sonic Browser as used for the experiment

The users estimated the data positions in the two dimensional scale without a comparison stimulus or a reference scale. Despite being pre-defined, i.e. being limited to the screen, the ranges of perceptual evaluations were relative to each user. All the users considered the perceptual space boundaries, as they reported at the end of the task, relative to their maximum value. In fact, we noticed an initial difficulty by the participants of referring to the screen space. They showed a preference of defining their own boundaries. In order to be able to compare the results of each participant, we decided to normalize the data.

## 5. Experimental Procedure

A workstation personal computer<sup>ii</sup> was used in the study. The experiment was conducted in the isolation room of the recording studio at the Computer Science Department at UL. The stimuli were presented by stereo headphones to the user through the Sonic Browser as shown in Figure 1, which allowed the users to listen to the sounds as many times as they wanted and to drag-and-drop them according to the dimensions on the axes. The users' speech and actions were recorded on video.

The general method of evaluation used in this study is the *Thinking-Aloud* method [8], where participants are asked to voice their thoughts while trying to accomplish the tasks, an example of a transcript taken using this method shown in Figure 2. The participants were introduced to the aims of the study and the application involved. They then familiarised themselves with the application for approximately five minutes. After familiarisation was completed they were asked to complete the tasks, and finally a debriefing. During the introduction, participants were shown the basic functions of the Sonic Browser.

Evaluator	Time	Task Comments	Task Duration
Observer	00:00:00	Start of Video	TASK x - MM:SS
Observer	00:03:18	Layout of Test Facilities	
Pilot 1	00:05:48	"I want to explore all the sounds before I begin"	
	01:25:00	"There I hear the force of the impact"	
	01:25:34	"I want to hear all the sounds before making a comment"	
	01:35:35	"This is a small object"	
	02:29:48	"I don't know if it is so high"	
	02:34:00	"This is a small object"	
	03:30:00	"Now I'm classifying about the size then I will think of the height"	
	03:35:50	"This is a small object, this is higher"	
	03:57:00	"This is a small object, this is higher"	
Observer	04:34:00	"You can decrease the size of the aura with the 'x' key or turn it off with the '10' keys"	
	04:41:00	ALPHA was turned off	
	05:03:00	"They are the same size but of different height"	
	05:19:00	"I will think ..."	
	06:28:18	"Now I'm adjusting the size"	
	06:33:33	"But this one I two I cannot recognise the size"	
	06:54:03	"The size and then I will think of the height"	
	07:12:20	"These two sounds have the same size but different heights. I will think about which one is higher"	
	07:44:30	"Also this one has the same size"	
	07:58:50	"And now they are right"	
	08:40:00	"It's difficult because I cannot imagine a scene"	
	09:10:40	"This one is higher than this one"	
	09:49:00	"It's the opposite." (Moved one sound then realised it was a sound which was dropped from a lower height)	
	11:11:23	"I'm trying to compare the different heights"	
	11:45:00	"This one is higher"	
	14:11:00	"The last two sounds have the same size but different heights. I cannot recognise which is higher"	
		"Now more or less I have adjusted everything but I now want to hear all the sounds and to adjust the size and scale of everything"	
	15:30:00	"More or less everything"	
	16:27:00	"Ok I'm happy"	
	16:48:00	"Ok I'm happy"	
	16:50:00	"Ahhh, there's an object here, last object" (Object partial obscured in the bottom right corner)	
	17:07:34	"Large, this is a small object"	
	17:40:15	"This is one in the middle of these two"	
	18:33:00	"I'm trying to move everything"	
	19:30:00	"I'm happy"	
	19:30:00	END TASK	19:21:02
Observer	19:40:00	"Could you select the file menu and select the second option 'save user report file'"	
Observer	19:50:00	"Could you type 'John' for the name?"	
Observer	20:04:00	"Thank you very much"	

SOUND SET for pilot one consisted of 5 sound object wooden bouncing balls and 5 recorded sounds of bouncing balls

Figure 2: Transcript taken using Think-Aloud Method

The debriefing was done immediately after the task. A questionnaire with semantic differentials was used for gathering the participants' reactions to the tasks and to the application.

## 6. Results

The results of the experiment with the layout of the individual sounds and their perceived realism are displayed by the square

symbol or by the x symbol as shown in Figure 3. The results of the participants perceived dimensions have been normalised to produce comparable graphs as shown in Figure 3 which present a generalised overview of each user's final judgements. The dimensions are, the x dimension as the size of the ball and the y dimension for the height of the fall.

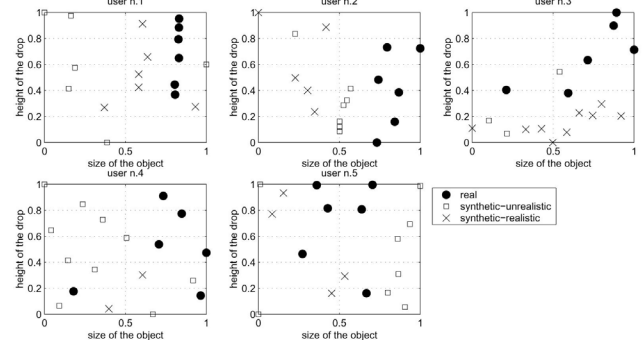


Figure 3: Graphical Representation of User's Perceived Scaling of Objects

The differences in classifications per sound are shown in Figure 4 for height differences and in Figure 5 for size differences. These differences per sound as shown by individual box plot's display the range of difference for each sound. The x-axis in Figure 4 represents the normalised height while the x-axis represents the size of the stimuli in Figure 5. Sounds with a larger visible box plot are those with the greatest differences between how individuals perceived their properties. Sounds with a very small or near invisible box plot are those sounds which individuals classified as having the same value within the particular dimension. These similarly classified sounds or stimuli with the smallest boxes represent sounds that convey the particular dimensions information well.

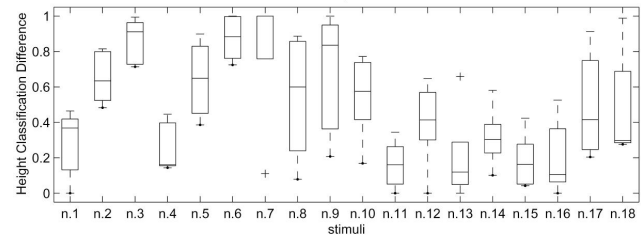


Figure 4: Box Plot Representation of Differences in Height Scaling, by stimuli

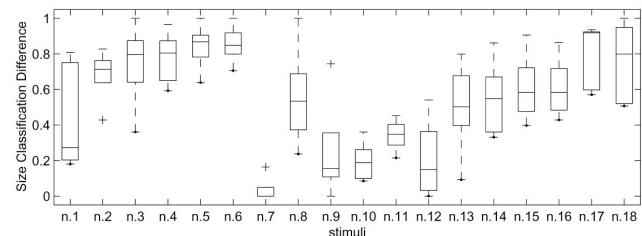


Figure 5: Box Plot Representation of Differences in Size Scaling, by stimuli

Examining the results we find that the sounds synthesized with only a single mode were well rated. The sounds in the main had both of their parameters estimated coherently by participants as shown by four of the six sounds (n.7, n.10, n.11 and n.14). The second approach which used sounds synthesised with two modes and a dropping event were not clearly estimated and provide a spread of answers across the participants. A possible explanation of this spread of estimations may be due to the presence of a "buzz tail" present at the end of these sounds, that conveys an unnatural impression of the event. The "buzz tail" is an effect present in the sound models where the model sound would keep 'bouncing' longer than a real sound would, a real sound would 'roll' on a surface after a number of bounces. The extra duration of 'bouncing' caused confusion for the participant's judging the sound.

The results of the realism task found amongst other results that none of the real sounds were classified as unrealistic. As this is the case, we'll focus on the synthesised sounds. There was no sound that was judged unrealistic by all participants. The most unrealistic sound was judged to be unrealistic by a consensus of three of the participants.

In the debriefing phase, a seven point Likert scale questionnaire with five sets of semantic differentials was filled out by the participants who were asked to express their opinions (from 0 to 6, where 0 is "poor" and 6 is "excellent"). In Figure 6, the results of the questionnaire with cumulative participant responses displayed per question can be seen.

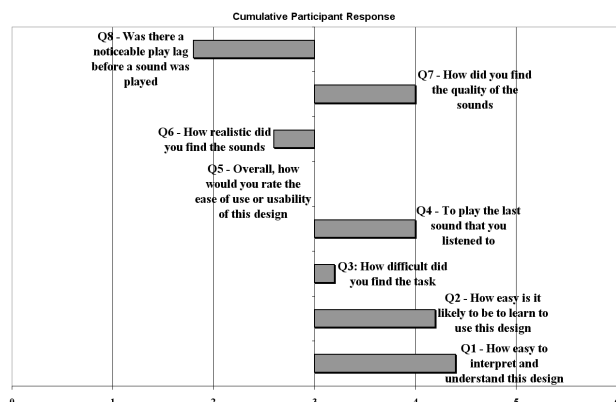


Figure 6: Results of questionnaire

Questions one and two were related to the learnability and interpretation of the application. Question three, which regards the participant's subjective difficulty in performing the task, found that participants perceived this be a non-trivial task. The result of question four, show that participants feel it was easy to replay the last sound played. Question five deals with the ease of use of the Sonic Browser for these tasks, which was found to have an average ease of use. Questions six and seven dealt with the realism and quality of the sounds. The results from these questions show that sounds were found to be of a high quality but that did not seem particularly realistic by the participants. This can be attributed to the inclusion of two different types of sound objects containing either one or two modes as well as the lack of room acoustics within the sound object sounds and the presence of a "buzz tail" at the end of the two mode sound object sounds. Questions eight concerned a slight delay when

playing an audio file with the Sonic Browser. The result of this question was a very noticeable delay when playing sounds within the Sonic Browser. The results of these questions showed that it was easy to learn and understand how to use the application for the task.

The *Thinking-Aloud* method returned several interesting results during the experiment. Many participants found a problem with the scales and on "deciding the scale whether to start with big or small sounds". One statement from several participants was that they perceived it was "much easier to judge size over height" in the scaling task. This can be seen in Figure 5 as the range of differences for scaling the size of objects are much smaller than for the related scaling ranges in judging the height of objects as shown in Figure 4, indicating that the sounds were more uniformly judged in the size scaling. The "use of speed of repetition as characteristic of height" was found to be helpful in classifying the height of a sound. Another common problem was that the "metallic zips distracts/confuses the classification of sounds", this refers to ending of each of the two mode sounds. Another issue illustrated by participants was that a "detailed comparison without reference points is very difficult and would be much easier with only a single scale" and this predicts the cognitive load of scaling the sounds a 2-dimensional space. The aura was found to be particular useful as "it allows me to see which is higher or which is lower by using pitch. The aura gives a comparison for similar sounds". Another important issue highlighted by participants was that the sound collection consisted of "three divisions (small, medium, large) and that it was very hard to compare between divisions but it was easy to compare within a division". The divisions refer to the three types of sounds within the space, real sounds, one mode sound objects and two mode sound objects. The participants also spoke about the different materials and surfaces as they found that the "different surfaces are very noticeable".

One participant (user n.3) performed the task in a very short period compared to the other participants but generated a less accurate scaling than other participants. This compared with other participants who found that "the longer I spent working with the sounds, the more difficult it was to sort them". This relates to a greater working knowledge of the sound collection and the difficulty in maintaining a consistent scale across multiple sounds. By concentrating on an initial reaction with a continuous exploration and classification of the sound collection it is possible to complete the scaling very quickly but the results showed that quality of the results were only of average quality compared to the other participants as shown in Figure 3. Participant debriefing revealed no sense of fatigue or boredom during the experiment due to the interactive nature of the display, which empowered the participant and allowed them feel in control of the experiment. The required time aspect was not covered within this experiment as compared to MDS as the MDS does not allow for a continuous overview and focus on the entire stimulus collection.

## 7. Discussion

There is a distinct difficulty in conveying information in more than one dimension, which is similar to the difficulty encountered with the parameterising of auditory icons [9, 10].

Synthesised sounds, even when found to be unrealistic, can still convey information in two dimensions. We find the realism of a sound does not affect its ability to convey information about particular perceptual dimensions. This result is similar to existing finding in this area regarding event perceptual of synthesised sounds [9, 11]. Another point is that with sound-to-sound comparison tasks such as psychoacoustic scaling, any minor temporal delay in playback is both obvious and irritating.

## 8. Future research

After examining the results of our experiment, there are a number of issues that have been illuminated. The most obvious is the removal of the delay with audio playback by creating a new audio layer or accessing the audio resources at a lower layer to improve playback performance. The second issue is to ensure that future Sound Object's synthesised sounds ensure that the "buzz tail" or prolonged bouncing at the end of the two mode sounds is removed to create more realistic sounds. The addition of features to form multiple levels of groups and associated subgroups of stimuli using borders and colouring and/or shading definitions with optional labels for each group or subgroup similar to the functionality of the Sonic Mapper [5] should be investigated as additional technique for the perceptual scaling task. A comparison between a MDS and this type of experiment should be conducted with a large subject base to compare this new technique with regards to classical MDS techniques.

## 9. Conclusions

The Sonic Browser offers a novel way of engaging users in psychoacoustic scaling experiments as well as reducing the boredom of these experiments for participants as well as providing results that are in the context of the whole stimuli set.

Our application allowed us to efficiently collect similarity data for a two-dimensional space with a relatively large numbers of sounds.

## 10. Acknowledgements

This work was supported by the European Union under the European Commission's Future & Emergent Technologies collaborative R&D programme (Sob – The Sounding Object Project – no IST-2000-25287; <http://www.soundobject.org>). We would like to express our thanks to all the postgraduate students at the University of Limerick who participated in this experiment.

## 11. References

- [1] E. Brazil, J. M. Fernström, G. Tzanetakis, and P. Cook, "Enhancing Sonic Browsing Using Audio Information Retrieval," presented at International Conference on Auditory Display ICAD-02, Kyoto, Japan, 2002.
- [2] E. Brazil, L. Ottaviani, and M. Fernström, "Experiments in a 2-D space using the Sonic Browser for validation and cataloguing of Sound Objects", In *Perceptual Assessment of models*, pp 40-76. SOb Project, 2002, Available at <http://www.soundobject.org>.
- [3] B. Arons, "A Review of the Cocktail Party Effect," *Journal of the American Voice I/O Society*, vol. 12, pp. 35-50, 1992.
- [4] B. Shneiderman, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 2 ed. Reading, MA, USA: Addison-Wesley, 1992.
- [5] G. P. Scavone, S. Lakatos, and C. R. Harbke, "The Sonic Mapper: An Interactive Program for Obtaining Similarity Ratings with Auditory Stimuli," presented at International Conference on Auditory Display ICAD-02, Kyoto, Japan, 2002.
- [6] M. Rath, "Sound design around a real-time impact model", In *Models and Algorithms for Sounding Objects*, pp 32-45. SOb Project, 2002, Available at <http://www.soundobject.org>.
- [7] R. Burro, "Interaction effects in bouncing ball", In *Physically-Based Sound Spaces and Psychophysical Scales*, pp 23-35. SOb Project, 2002, Available at <http://www.soundobject.org>.
- [8] C. Lewis, "Using the "Thinking-aloud" Method in Cognitive Interface Design," IBM, Yorktown Heights, Research Report RC 9265, 1982.
- [9] W. W. Gaver, "Synthesizing Auditory Icons," presented at Interchi'93, 1993.
- [10] W. W. Gaver, "Using and Creating Auditory Icons," in *Auditory Display: Sonification, Audification and Auditory interfaces*, G. Kramer, Ed. Reading, MA, USA: Addison-Wesley Publishing Company, 1994, pp. 417-446.
- [11] W. Warren and R. Verbrugge, "Auditory perception of breaking and bouncing events: A case study in ecological acoustics," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 10, pp. 704-712, 1984.

---

<sup>i</sup> An installer for the application and instructions for its installation are available at the Sounding Object Project (SOB) website. The Sounding Object Project (SOB) homepage is at <http://www.soundingobject.org>

<sup>ii</sup> AMD Dual Athlon MP, 1500MHz, 1024 MB RAM, 19" display with 1280 x 1024 pixels at 85 hertz in 32-bit colour, Creative Labs SoundBlaster Live! Platinum sound card, stereo headphones Sony MDR-CD280, Microsoft Windows 2000 Professional v5.0.2195 Service Pack 3.